

## 结合运动信息与表观特征的行人检测方法\*

郭立君<sup>1,2,3+</sup>, 刘曦<sup>2,3</sup>, 赵杰煜<sup>1,2</sup>, 史忠植<sup>2</sup>

<sup>1</sup>(宁波大学 信息科学与工程学院, 浙江 宁波 315211)

<sup>2</sup>(中国科学院 计算技术研究所, 北京 100190)

<sup>3</sup>(中国科学院 研究生院, 北京 100049)

### Pedestrian Detection Method of Integrated Motion Information and Appearance Features

GUO Li-Jun<sup>1,2,3+</sup>, LIU Xi<sup>2,3</sup>, ZHAO Jie-Yu<sup>1,2</sup>, SHI Zhong-Zhi<sup>2</sup>

<sup>1</sup>(College of Information Science and Engineering, Ningbo University, Ningbo 315211, China)

<sup>2</sup>(Institute of Computer Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>3</sup>(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: E-mail: guolijun@nbu.edu.cn, http://icst.nbu.edu.cn/members/guolijun.html

**Guo LJ, Liu X, Zhao JY, Shi ZZ. Pedestrian detection method of integrated motion information and appearance features. Journal of Software, 2012, 23(2): 299-309.** <http://www.jos.org.cn/1000-9825/4030.htm>

**Abstract:** This paper proposes a method of pedestrian detection that takes both motion information and appearance features into account. This could be done by integrating motion information into the segmentation algorithm in the framework, which performs the validation of segmentation on candidate detection windows obtained by the appearance detector. The paper considers that better segmentation results can raise the detection accuracy. Shape features are obtained by integrating color information indirectly into pedestrian detection by using motion information to model foreground/background distribution of moving object. Better detection performance benefits from the complementary advantages between shape features and pedestrian appearance detector. The claim is supported by these experiments based on CAVIAR and the test video with pedestrians.

**Key words:** pedestrian detection; shape prior; motion information; graph cut

**摘要:** 提出一种结合运动信息与表观特征的行人检测方法. 在对通过表观检测子获得的候选检测窗口执行分割验证的框架中, 将运动信息融入到基于图像序列的对象分割算法中, 通过获取更准确的分割结果来提高对候选检测窗口的检测准确率. 该方法利用运动信息更新运动对象的前景/背景分布模型, 将颜色信息间接地融入行人检测中, 并通过形状特征表现出来, 与行人表观检测子形成互补的特性, 获得更好的检测结果. 上述结论在 CAVIAR 视频以及行人检测视频中得到了实验验证.

**关键词:** 行人检测; 形状先验; 运动信息; 图割

中图法分类号: TP318 文献标识码: A

\* 基金项目: 国家自然科学基金(61175026, 60903141); 国家重点基础研究发展计划(973)(2007CB311004); 浙江省自然科学基金(D1080807); 浙江省新一代移动互联网用户端软件科技创新团队项目(2010R50009); 宁波市自然科学基金(2011A610193)

收稿时间: 2010-07-07; 定稿时间: 2011-03-07

行人检测在视觉监控、汽车自动驾驶、机器人视觉、基于内容的图像与视频索引以及高级人机交互等方面有广泛的应用,吸引了众多研究人员的注意.在实际应用中,由于场景的复杂性、视角与尺度的变化、人体姿态与着装的多样性以及部分遮挡等因素,使得行人检测具有极大的挑战性.近几年来,随着研究的深入,研究人员提出了各种人体检测方法,包括基于梯度方向的 HOG(histogram of oriented gradient)检测子<sup>[1]</sup>、类似形状的检测子<sup>[2]</sup>、基于部件的检测方法<sup>[3,4]</sup>、多种特征融合的行人检测方法<sup>[5,6]</sup>以及结合各种上下文(context)信息的检测方法<sup>[7]</sup>等.上述方法都是基于图像的静态人体检测方法,尽管取得了不错的检测效果,然而在减少误检(false positive)、检测定位准确率以及检测效率等方面仍然需要进一步提高.

心理学与神经科学方面的实验都表明,运动特征是感知行人的一个重要线索.较早利用运动信息的行人检测方法都是通过在多帧图像中跟踪运动对象并分析运动的周期性特征来判别人.这类方法对行人的检测明显要滞后,并且无法检测到静止或运动缓慢的行人.为了能够同时准确地检测运动人体与静止人体,结合运动信息与表观特征是一种有效的行人检测方法.这类方法的关键是找到一种合适的结合方式,充分挖掘两种特征在行人检测中的判别特性,进一步提高行人检测准确率.受基于分割-验证检测方法的启发<sup>[7,8]</sup>,本文提出了一种结合运动信息与表观特征的行人检测方法.该方法首先通过基于表观特征的人体检测子<sup>[9]</sup>获取当前帧图像中候选的人体检测窗口;然后,以离线训练获得的人体对象的全局形状特征为先验,通过基于图像序列的对象分割获取候选检测区域的分割掩模,在分割算法中融入运动信息以提高分割的准确性;最后,通过对对象分割掩模验证当前候选检测结果,消除基于表观检测子的误检,进而提高整个算法的检测准确率.

## 1 相关的研究工作

基于图像的人体检测研究已经进行了很长时间<sup>[10-12]</sup>,近几年的研究在检测效果上取得了显著的进步.Dalal 与 Triggs 等人<sup>[1]</sup>提取梯度方向直方图(HOG)特征,采用多尺度窗口扫描结合 SVM 分类器的方式达到了较好的人体检测效果.Felzenszwalb 等人<sup>[4]</sup>将多尺度的变形部件模型与 HOG 特征相结合,提出一种隐式 SVM 判别方法来解决部件位置为隐变量的训练问题,进一步提高了行人检测准确率.Dollar 等人<sup>[3]</sup>同样提出了一种基于部件模型的行人检测方法,但不同的是,利用小波特征与 Boosting 方法训练部件模型.Sabzmeydani 等人<sup>[2]</sup>通过 AdaBoost 学习由底层梯度信息组合形成的类似形状的中层特征,并进一步利用 AdaBoost 选择形状特征子集进行人体检测,获得了超过标准 HOG 检测准确率的检测效果.Tran 等人<sup>[13]</sup>利用梯度的局部直方图与梯度的局部 PCA 特征学习人体模型完成人体检测.Wu 等人<sup>[14]</sup>在 Boosting 框架与 edgelet 特征相结合的基础上,通过自动构造层次树解决多视角的行人检测问题.

为了进一步提高人体检测的准确率,研究人员提出了基于上下文的静态人体检测以及多特征融合的行人检测方法.基于上下文的对象检测一般都利用基于表观特征的检测子以窗口扫描的方式获得大量的候选检测窗口,再通过各种上下文信息对候选检测窗口进行验证,以此提高检测的准确率.Divvala 等人<sup>[7]</sup>利用文献[4]中提出的检测子获取特定对象(包括人体)检测的候选窗口,基于各种可能的 Context 特征训练若干分类器,并通过逻辑回归方式将多个分类器合并在一起,完成对候选检测窗口中对象的出现、定位以及尺寸的判别,并结合分割方法获取对象的空间支持.Ramanan 等人<sup>[8]</sup>同样在通过基本检测子获得候选人体检测窗口后,通过分割获取窗口中对象的形状特征,并基于该形状训练人体分类器完成对检测窗口的验证.

行人检测方法的另一个研究方向是多特征融合.Wojek 等人<sup>[5]</sup>评价了多个人体检测特征组合的检测效果.Wu 等人<sup>[6]</sup>提出了一种兼顾检测效率与检测准确性的多特征结合的人体检测方法,分别基于 HOG、edgelet 以及协方差等特征训练多个分类器,自动学习分类器的分类性能,并据此建立分层次的多级人体检测分类器.在与运动特征的结合方面,Jones 等人<sup>[15]</sup>与 Viola 等人<sup>[11]</sup>提出了结合表观特征与运动特征的行人检测方法,分类器的输入来自连续的两帧或多帧图像中同一小窗口内的图像,基于连续两帧图像的帧差异以及两帧图像的方向偏移提取运动特征,基于单独一帧图像提取 Haar 表观特征,最终基于上述两类特征训练 Adaboost 分类器进行行人检测.这种运动信息的结合方式增强了行人检测特征的判别特性.然而,文献[15]与文献[11]中的运动特征提取只能限制在固定场景下.此外,由于要求在连续多帧图像的检测窗口内,行人不能运动到检测窗口之外,而检测

窗口在多帧图像序列中是有固定位置和大小,因此只能完成针对很小尺寸的行人检测,限制了行人检测的应用环境.Tian 等人<sup>[16]</sup>首先利用表观检测子(基于类似 Haar 特征的 Adaboost 分类器)初步判别单帧图像中的行人检测框,根据连续帧中检测框的颜色与空间信息的相似性判别一个检测框是否多次出现,并将多次出现的检测框作为真实的行人检测结果,以此消除误检.这种基于颜色与空间信息的时序分析方法(间接地利用了运动信息)可以消除一部分在表观检测子中偶尔出现的错误检测框,却无法消除在表观检测子中连续出现的误检,因此并没有真正弥补相应的表观检测子在行人检测能力上的不足.Dalal 等人<sup>[17]</sup>与 Wojek 等人<sup>[18]</sup>都通过光流提取运动特征,并与一种或多种表观特征相结合进行行人检测.文献[17]中的检测结果没有显著超过他们自己的静态 HOG 检测子的检测准确率<sup>[19]</sup>,文献[18]结合运动特征的检测结果对摄像机与光照条件特别敏感.

由上述分析可知,由于无法同时提取多种有关人体检测的判别特性,一种静态的人体检测子无法获得最佳的检测效果.通过多种特征融合或借助于各种上下文信息能够有效提高检测的准确率.其中,运动特征由于蕴含着丰富的行人检测信息而受到重视.现有的结合表观特征与运动信息的检测方法都存在一些缺陷,其原因是缺乏合适的融合方式,使得这种融合能够充分挖掘两种特征各自的判别特性,互相弥补各自的不足.本文提出了一种结合运动信息与表观特征的检测方法.对比实验表明,该方法在降低误检率、提高检测准确率方面达到了更好的检测效果.

## 2 结合运动信息的行人检测

文献[7,8,20]研究表明,通过分割获得的各种线索可以有效地解决基于窗口扫描方式的人体检测子产生的误检问题,进而提高检测的准确率.为了解决由于分割动作而产生的计算可行性问题,本文采用与文献[7,8]类似的“假设+验证”的检测框架,即只针对通过 HOG 检测子获得的候选检测窗口执行分割验证.为了获得与 HOG 具有互补特性的分割线索,文中通过分割获取候选检测窗口前景对象(人体)的全局形状特征(分割掩模),并联合该特征与 HOG 检测结果共同训练人体分类器对候选检测窗口进行验证.在连续多帧图像中,基于行人图像(前景)变化不大,但背景图像会发生变化的考虑,本文将运动信息引入到对象分割算法中,在获得更准确分割结果的同时,提高了行人检测的准确率.具体的检测过程如图 1 所示.

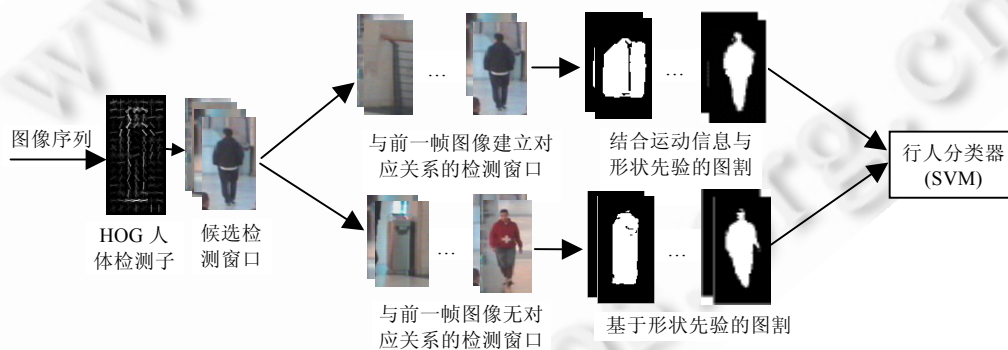


Fig.1 Pedestrians detection framework with integrated motion information

图 1 结合运动信息的行人检测框架

1. 针对当前帧图像,利用 HOG 检测子获得候选检测窗口(适当降低检测的阈值设置,以便能够提取更多的候选检测窗口).
2. 通过质心与长宽比确定当前帧每个候选检测窗口与前一帧中检测窗口的对应关系,并依据对应结果确定不同策略的分割方法,获取检测窗口区域内的对象分割掩模.
  - 若当前候选检测窗口与前一帧中的检测窗口没有对应,则认为该检测窗口包含新出现的候选检测对象.此时,采用基于形状先验的图割方法.形状先验的获取及相应图割算法的详细过程见第 2.1 节.
  - 若当前候选检测窗口与前一帧中的某检测窗口有对应关系,则认为当前窗口与前一帧对应的检测

窗口中包含有同一个候选检测对象.此时,利用结合运动信息与形状先验的图割方法,具体的图割方法见第 2.2 节.

通过分割获取的对象掩模将作为当前检测窗口的前景模型保存,以备后面对应的窗口能够基于该模型修正分割结果,这也是结合运动信息的本质所在.

3. 最终,将候选检测窗口的分割结果(二值掩模)与当前检测窗口的 HOG 检测得分一起送入行人分类器,验证当前候选检测窗口是正确的还是误检,离线训练行人分类器的过程见第 2.3 节.

在上述分割过程中,我们将运动信息融合在基于多帧图像的图割过程中,以便能够获得更准确的对象分割结果,进而对候选检测窗口给出准确的检测结果.如果在连续多帧图像中获得同一运动对象的多个候选检测窗,则认为这些检测窗口中的运动对象(前景)具有相同的颜色分布,而这些窗口中的背景图像存在差异,具有各自独立的颜色分布.因此,在当前帧检测窗口的对象分割过程中,可以融合前面多帧中对应检测窗口的分布模型建立当前检测窗口的前景/背景模型,提高分割的准确性.

上述基于多帧图像的对象分割思想一般需要在准确跟踪的前提下才能获得较好的分割结果.本文的重点不是运动对象的跟踪,而是通过获得满意的窗口中对象分割来验证检测的准确性.因此,文中没有采用在运动对象检测基础上的卡尔曼滤波方法或粒子滤波方法执行多运动目标跟踪,而是通过简单地计算检测窗口的质心与长宽比建立连续两帧中检测窗口的对应关系(考虑到候选检测窗口是通过多尺度窗口扫描方式获得的,这里在质心的基础上又增加了长宽比因素).研究表明,在中等分辨率以上的环境中(人体高度大于 40 个像素),HOG 检测子对人体检测具有较强的稳定性,能够检测到大多数的人体对象.当然,这同时也导致数量众多的误检窗口.因此,本文方法的目标就是通过降低误检率来提高检测准确率.正是 HOG 对人体检测的稳定性,使得我们提出的基于质心与长宽比的方法能够获得较好的跟踪效果.即使是在多个行人距离较近的环境中,这种简单的方法也能在两帧的检测窗口间建立正确的对应关系.此外,文中把没有建立对应关系的检测窗口(包括由于跟踪失败导致的情况)视作新出现的检测窗口,并利用自顶向下的先验知识分割窗口中的对象,即采用基于形状先验的图割方法,仍能达到很好的验证效果.因此,本文中的结合运动信息的分割验证方法对跟踪具有较强的鲁棒性.

## 2.1 基于形状先验的分割算法

在结合运动信息的行人检测框架中,根据当前候选检测窗口是否与前一帧图像的检测窗口建立对应关系,分别采用两种不同策略的分割算法获取当前候选检测窗口的分割掩模:一种是基于形状先验的分割;另一种是结合运动信息与形状先验的分割.两种策略的分割都建立在基于能量最小的图割方法<sup>[21,22]</sup>上,只是能量函数中一元项的计算方法不同,目的是尽可能多地利用判别信息获得准确的分割结果.

根据 MRF-MAP 框架理论<sup>[22]</sup>及其在图像分割中的应用,二值图割是一种使得下面的能量函数取值最小的图割方法:

$$E(f) = \sum_{x \in X} V_x(f_x) + \sum_{\{x,y\} \in N} V_{\{x,y\}}(f_x, f_y) \quad (1)$$

其中,  $f$  对应一种分割结果,  $X$  是图像的像素集合,  $f_x \in \{fg, bg\}$  是像素  $x$  的分割标签,  $N$  是一个像素的邻域,文中取四邻域.能量函数  $E(f)$  的第 1 项称为一元项或似然项,定义了单独一个像素隶属于前景或背景的程度;第 2 项称为二元项或先验项,定义了像素在邻域范围内的平滑程度.由于分割中采用的图模型不同,这两项能量的具体形式也不相同.

根据前面检测框架所述,针对与前一帧没有对应关系的候选检测窗口,我们将离线获得的人体形状先验信息融入到二值图割算法中.通常情况下,形状先验以显式方式作用于能量函数的一元项中,如 Rihan 等人<sup>[23]</sup>以一种形状能量(到对象边界的距离函数)的方式结合到一元项中.显式方式倾向于将候选检测窗口强制分割成具有类似人体形状的对象.为了能够最大程度地借助于图像低层特征获得分割结果,并通过后面的形状分类器验证分割结果是否为行人,我们采用了类似于 Ramanan<sup>[8]</sup>中隐式结合形状先验的方法.本文中,能量函数的一元项通过基于窗口内颜色直方图建立前景/背景的概率分布模型来计算.为了结合形状先验信息,我们将检测窗口规范化到与获取人体形状先验的掩模窗口一致的尺寸上,以便通过形状先验确定检测窗口中每个像素属于前景的概率

$p_{fg}(x)$ .在此基础上,根据下面的公式计算检测窗口中前景/背景颜色的直方图分布模型:

$$\Pr(fg(k)) \propto \sum_x p_{fg}(x)[h(x) = k] \tag{2}$$

$$\Pr(bg(k)) \propto \sum_x (1 - p_{fg}(x))[h(x) = k] \tag{3}$$

其中, $k$ 为直方图 Bin 的索引.因为无须考虑光照的影响,我们直接采用了 RGB 颜色模型.针对  $R,G,B$  这 3 个轴各自设置 16 个 Bin, $k \in \{1,2,\dots,4096\}$ .分割算法中只考虑非零的 Bin,一个检测窗口内,非零的直方图 Bin 的数量一般只有几百个. $h(x)$ 为像素  $x$  所属的 Bin 的序号, $[\cdot]$ 是布尔运算.根据 MRF-MAP 框架理论,一元项采用概率分布密度的负对数来估计:

$$V_x(f_x=fg) = -\log(\Pr(fg(h(x)))) \tag{4}$$

$$V_x(f_x=bg) = -\log(\Pr(bg(h(x)))) \tag{5}$$

对于能量函数中的二元项,文中采用类似文献[21]的方法定义为

$$V_{\{x,y\}}(f_x, f_y) = \exp\left(-\frac{\|v_x - v_y\|^2}{2\sigma^2}\right) [f_x \neq f_y] \tag{6}$$

其中, $[\cdot]$ 是布尔运算, $v_x$ 为像素  $x$  的颜色向量, $\sigma$ 定义为所有相邻像素颜色距离的平均值.在确定了能量函数(1)中的两项组成形式后,基于能量函数最小的分割结果,可以通过最大流-最小割(min-cut max-flow)<sup>[21]</sup>方法有效地得到.

人体形状先验:前面用到的人体形状先验是指在同一坐标系下、相同尺度的窗口中,独立统计每个像素属于前景(人体)的概率  $p_{fg}(x)$ 以及属于背景(环境信息)的概率  $1-p_{fg}(x)$ .建立这样的概率分布模型需要提供一组基于窗口的人体分割的正样本.给定这样的一组样本  $T$ ,每个样本中的分割结果表示为  $f^t$ ,我们可以采用最大似然估计方法计算  $p_{fg}(x)$ :

$$p_{fg}(x) \stackrel{MLE}{=} \frac{1}{T} \sum_t f_x^t \tag{7}$$

获得足够多的准确分割样本是件繁琐的工作.我们采用了一种半监督的学习方法建立人体形状先验模型.针对矩形框标注的人体库(文中选择 INRIA 的人体训练库),首先将人体标注的矩形框适当地扩大(扩大的比例与 HOG 检测子中人体边界的比例相同),以便包含人体周围的背景信息.首先根据下面的公式计算人体边框先验:

$$p_{fg}^0(x) = [x \in \text{标注矩形框}] \tag{8}$$

人体边框先验相当于将初始人体标注的矩形框内像素的前景概率为 1,外扩区域内(人体周围的环境信息)像素的前景概率为 0.将上述人体边框先验作为初始的形状先验,采用本节前面提到的能量函数最小化的方法分割当前的训练样本.将分割结果归一化到标准尺度的窗口大小后(实验中 HOG 检测子定义的窗口大小)代入公式(7)中计算最终的人体形状先验,如图 2 所示.这个过程离线执行,获得的人体形状先验用于前面提到的基于形状先验的分割.

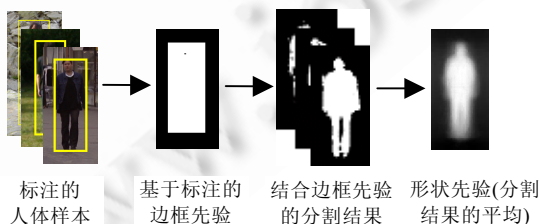


Fig.2 Obtaining of human shape prior

图 2 人体形状先验的获取

## 2.2 结合运动信息的分割算法

在行人检测框架中,若当前候选检测窗口与前一帧的检测窗口建立了对应关系,则意味着两帧或者多帧中具有对应关系的检测窗口中包含同一个人体对象.此时可以假定多帧检测窗口内像素的背景模型是独立的,前景模型是相同的.因此,在对当前候选检测窗口的分割过程中,可以同时引入运动信息与人体形状先验.具体分割算法如下:

- (1) 利用人体形状先验计算当前检测窗口中前景分布模型  $M_{fg\_new}$  与背景分布模型  $M_{bg}$ ,即通过公式(2)与公式(3)计算前景/背景颜色的直方图分布模型.
- (2) 结合前面若干帧的前景模型修正当前检测窗口的前景分割模型:

$$M_{fg} = (1 - \alpha)M_{fg\_new} + \alpha M_{fg\_old} \quad (9)$$

其中,  $M_{fg\_old}$  是前一帧对应窗口的前景模型;  $\alpha \in [0.5, 1]$  为比例系数,用于调整  $M_{fg\_new}$  与  $M_{fg\_old}$  在形成最终的前景模型  $M_{fg}$  中的作用.

- (3) 在新的前景/背景模型基础上,计算公式(1)中的一元项与二元项,执行基于能量函数最小的图割算法.

这种结合运动信息的图割方法在以下两个方面提高运动对象分割的准确性,进而通过验证方式提高候选人体检测的准确性:一方面,在一定程度上消除由于多个行人连在一起或与其他物体交叠对分割产生的不利影响;另一方面,在对当前候选窗口分割过程中,降低人体形状先验对非直立状态人体的不利影响.我们认为,  $M_{fg\_old}$  是前面已经确定的人体前景模型,即通过前面的检测结果反过来证实了  $M_{fg\_old}$  作为人体前景模型的合理性与准确性.这种情况下,将该模型作用于同一个人体后续检测窗口的前景模型,显然可以提高分割的准确性.另外,由于  $M_{fg\_new}$  主要依据人体形状先验获得,更多地体现了人体形状先验的作用,  $M_{fg\_old}$  则更多地体现了运动信息的利用.公式(9)中  $\alpha$  的取值在 0.5~1 之间,初值取 0.5,每次递增(本文中递增值为 0.1),直到取值为 1 后保持不变.针对重新建立对应关系的检测窗口,  $\alpha$  重新取初值.  $\alpha$  的这种取值方式,目的是在建立对应关系的窗口分割中逐步增强  $M_{fg\_old}$  的作用,削弱  $M_{fg\_new}$  的作用.即增强运动信息的作用,削弱形状先验的作用.上述结合运动信息的方式对分割准确性的影响在第 3 节的实验中得到了验证(详见图 5).

## 2.3 行人检测分类器

获得候选检测窗口的分割掩模以后,最终将掩模与窗口的 HOG 检测得分一起送入行人分类器,给出检测结果.行人分类器通过离线训练获得,训练的正负样本分别来自人体窗口与非人体窗口的分割结果.首先,通过 HOG 人体检测子从 INRIA 的人体图像库的训练集中提取候选的检测窗口.根据训练集中的标注信息,我们可以确定这些检测窗口哪些是包含人体的,哪些是非人体的检测窗口;然后,利用基于形状先验的分割方法分割这些检测窗口,将人体窗口的分割结果作为正样本,非人体窗口的分割结果作为负样本,由此获得一个训练集合.我们将分割的掩模结果与 HOG 的检测得分定义为检测窗口的特征,即

$$F_{window} = \{f_1, f_2, \dots, f_k, score\} \quad (10)$$

根据上面获得的训练集训练一个线性 SVM 分类器,对人体检测窗口与非人体检测窗口进行分类.

## 3 实验与分析

测试视频:目前用于行人检测的视频图像库中,最全面的应该是 TUD-Brussels<sup>[18]</sup> 视频序列库.该视频序列是基于运动平台获取的行人视频库,图像情景复杂,含有丰富的行人内容.然而该视频库并不适合本文中结合运动信息的行人检测算法,原因是,该视频库中的图像序列都是一对一对出现的,即一对图像是连续的两帧图像,而两对序列之间并不连续.这对于获取类似光流的运动特征是足够的,但对于本文中将多帧序列的运动信息融入对象分割中的方法略显不足.本文选用了 TUD-Brussels 之外的两段视频进行测试.一段来自 CAVIAR 项目具有固定背景的视频序列<sup>[24]</sup>,CAVIAR 是关于动作识别的一个项目,提供了各种环境下的人体运动视频.我们从中选取一段商店走廊的视频,这段视频含有相对较多的行人,且背景环境较为复杂(参见图 4(a)、图 4(b)).另一段测试视频则是我们自己通过运动中摄像机获取的一段视频(参见图 4(c)、图 4(d)).



我们在实验中对比分析了 3 种行人检测算法在两段测试视频中的检测效果.3 种算法分别为标准 HOG 静态人体检测算法、基于形状先验的分割验证算法(HOG+shape-segmentation,简称为 Hog+S-Seg)以及同时结合运动信息与形状先验的分割验证算法(HOG+motion and shape-segmentation,简称为 Hog+MS-Seg).标准 HOG 检测算法使用了提供开放代码的 fastHOG 检测子<sup>[9]</sup>,后面两种方法都是在标准 HOG 检测算法基础上实现的:首先,通过降低标准 HOG 检测子的得分阈值获得更多的候选检测窗口(返回窗口的坐标、检测尺度以及检测得分);然后,分别利用不同的分割验证算法完成行人检测.针对我们实验中选用的两个视频库,根据分辨率的不同,每帧通过基本检测子确定的候选检测窗口规模从十几个到几十个不等.这些候选检测窗口是执行了基于均值漂移算法的非最大抑制的结果,即合并了同一检测对象在不同尺度以及相近位置上全部的检测窗口.文中用到的人体形状先验以及行人检测分类器都基于 INRIA 的人体图像训练库学习获得,学习过程中将窗口都规范化成与 HOG 检测子相同的窗口大小.

文中采用了类似 PACSAL VOC 中的评价方法,即计算基于准确率与召回率曲线(PR-curve)的平均准确率(AP).基于对检测结果评价合理性的考虑,对测试的每段视频随机确定从某个位置开始,每隔 5 帧给出一幅图像对应 3 种检测算法的检测结果.每段视频各提取 200 帧图像作为检测结果,绘制各种算法的准确率与召回率,并最终计算出平均准确率.由于上述候选检测窗口都是基于基本 HOG 检测子获得的,基本 HOG 检测子选定的检测窗口为  $64 \times 128$ ,针对分辨率较低的人体对象很难检测到.因此,在评价不同算法的检测结果时,本文只考虑高度大于 40 的人体对象(对应的检测窗口高度一般大于 60),这并不影响算法之间的可比性.图 3 给出了 3 种检测算法在 CAVIAR 与我们自己测试视频上的 PR 曲线及平均准确率.

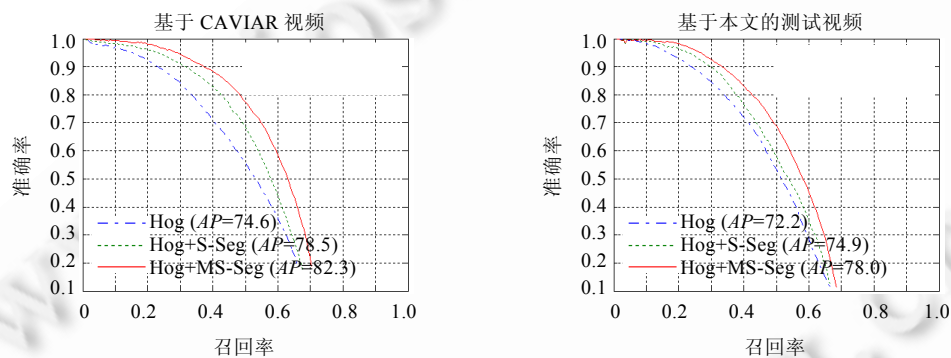


Fig.3 PR curves of three detection algorithms on two test videos

图 3 3 种检测算法在两个测试视频上的 PR 曲线

从图 3 可以看出,在两段测试视频上,两种基于分割验证的检测算法相对于基本检测子的检测准确率都有提高.其中,基于形状先验的分割验证算法在两段视频中分别提高了 3.9 和 2.7 个百分点,而利用运动信息的检测方法分别提高了 7.7 和 5.8 个百分点.分析原因是:两种基于分割验证的行人检测算法在 HOG 特征之外还利用了颜色信息;HOG 检测子是一种基于梯度方向的特征,只利用了图像的灰度信息.由于人体着装以及周围环境颜色的多样性,颜色信息并不能直接作为人体检测的线索.然而在基于分割验证的行人检测方法中,利用颜色信息建立分割过程中的前景/背景分布模型,进而获取对象的形状特征,并结合形状特征与 HOG 特征共同完成行人检测.丰富了行人检线索,提高行人检测的准确率也就成为必然.

两种分割验证的检测算法中,同时结合运动信息与形状先验的分割验证算法优于只利用形状先验的分割验证算法的检测效果.在两段测试视频中,前一种方法的平均检测率比后一种分别提高了 3.8 和 3.1 个百分点.其原因是,相对于后一种算法,前一种算法额外利用了运动信息获取更准确的形状特征.这一点在两段测试视频中都明显地体现了出来,尤其是针对具有更多的行人数量,行人出现的情况也相对更复杂,运动信息的特性体现得更充分一些.图 4 给出了 3 种检测算法在两段测试视频中部分帧上的检测结果.图 5 通过分割结果说明相对

于只利用形状先验的分割验证方法,结合运动信息的方法能够提高运动对象分割的准确性,并能在一定程度上消除由于多个行人连在一起或背景复杂对分割的不利影响.



Fig.4 Detection result of three detection algorithms on partial frames of two test videos

图 4 3 种检测算法在两段测试视频中部分帧上的检测结果

图 4 中共取 4 帧图像(对应 4 行),图 4(a)、图 4(b)来自 CAVIAR 视频,图 4(c)、图 4(d)来自本文的视频.每帧的 3 个检测结果(从左至右)分别对应 Hog,Hog+S-Seg 以及 Hog+MS-Seg 对该帧图像的检测结果.图像中,浅色框为正确检测结果,深色框表示误检结果,而黑色框表示由于分割结果不准确造成的漏检.可以看出,分割验证算法能够在提高召回率的情况下提高检测的准确率(对应着更多的候选检测窗口以及对误检窗口的过滤).但也有少数正确的候选检测窗口由于分割结果不准确而被过滤掉了,如图中黑色框标记,其中一个原因是尺寸较小



窗口(分辨率较低)的分割结果不准确.图5给出了3个人体候选检测窗口的分割实例.可以看出,由于背景以及行人连在一起等因素的影响,使得基于形状先验的分割验证方法将这样的候选检测窗口过滤掉,无法给出正确的检测结果.在通过结合运动信息与形状先验的分割验证方法中,明显改善了分割效果,因此仍能正确判别这样的候选检测窗口.



Fig.5 Different effects of the two segmentation validation methods on detection results

图5 两种分割验证方法对检测结果的影响

在同时结合运动信息与形状先验的检测实验中,有 65%的包含人体的候选窗口最终通过结合运动信息的算法给出的结果.这说明有一多半包含人体的候选检测窗口都建立了与前一帧的对应关系.与没有建立对应关系的候选检测窗口相比,结合运动信息与形状先验的分割算法相对于只利用形状先验的分割算法多了两项运算,分别是对应关系的计算以及前景模型的计算.对应关系通过简单的质心距离与长宽比来确定,因此计算量取决于候选检测窗口的数量;而前景模型的合并只是简单的向量相加.因此,在候选检测窗口数量不多的情况下,通过结合运动信息与形状先验的分割算法进行行人检测,在运算量没有显著增加的情况下,检测的准确率可以得到较大的提高.对于候选检测窗口数量众多的情况,也可以采用首先对候选检测窗口执行基于形状先验的分割验证,对于验证没有通过的窗口,再采用结合运动信息与形状先验的分割验证算法.

#### 4 结束语

运动信息中蕴含着丰富的行人判别信息,如何结合运动信息完成行人检测,一直是行人检测的一个研究方向.本文提出一种结合运动信息与表观特征的行人检测方法,在对基本表观检测子获得的候选检测窗口执行分割验证的框架中,将运动信息融入到分割算法中,通过获取更准确的分割结果来提高对候选检测窗口的检测准确率.这种融合运动信息的方法,利用运动信息更新运动对象的前景/背景分布模型,将颜色信息间接地融入行人检测中,并通过形状特征表现出来,与行人表观检测子形成互补的特性,获得更好的检测结果.

在接下来的工作中,我们将继续在以下两个方面展开进一步的研究:(1) 算法的运算效率问题.本文中算法的效率体现在两个方面:一是基于多尺度窗口扫描方式的 HOG 人体检测子,利用 CUDA 并行架构实现的 FastHog 已经达到实时检测的效果;二是基于窗口的对象分割,目前已有研究提出了比图割更高效的分割算法,

且取得了不错的分割效果。(2) 此外,建立在行人检测基础上的行人识别(person reidentification)是我们下一步研究的重点。

**致谢** 本文实验中的 HOG 检测算法是通过调用 FastHog 提供的库函数来实现的,部分测试视频来自 CAVIAR 视频库。在此,对 FastHog 函数库源代码的提供者 Victor Adrian Prisacariu 以及 CAVIAR 项目组表示感谢。

#### References:

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. San Diego, 2005. 886–893. [doi: 10.1109/CVPR.2005.177]
- [2] Sabzmeydani P, Mori G. Detecting pedestrians by learning shapelet features. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Minneapolis, 2007. 1–8. [doi: 10.1109/CVPR.2007.383134]
- [3] Dollár P, Babenko B, Belongie S, Perona P, Tu ZW. Multiple component learning for object detection. In: Proc. of the 10th European Conf. on Computer Vision. 2008. 211–224. [doi: 10.1007/978-3-540-88688-4\_16]
- [4] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Anchorage, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587597]
- [5] Wojek C, Schiele B. A performance evaluation of single and multi-feature people detection. In: Proc. of the 30th DAGM Symp. on Pattern Recognition. Munich, 2008. 82–91. [doi: 10.1007/978-3-540-69321-5\_9]
- [6] Wu B, Nevatia R. Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Anchorage, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587749]
- [7] Divvala SK, Hoiem D, Hays JH, Efros AA, Hebert M. An empirical study of context in object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2009. 1271–1278.
- [8] Ramanan D. Using segmentation to verify object hypotheses. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. Minneapolis, 2007. 1–8. [doi: 10.1109/CVPR.2007.383271]
- [9] Prisacariu VA, Reid ID. FastHOG—A real-time GPU implementation of HOG. Technical Report, 2310/09, OUEL, 2009.
- [10] Ioffe S, Forsyth DA. Probabilistic methods for finding people. Int'l Journal of Computer Vision, 2001,43(1):45–68. [doi: 10.1023/A:1011179004708]
- [11] Viola P, Jones MJ, Snow D. Detecting pedestrians using patterns of motion and appearance. In: Proc. of the 9th IEEE Int'l Conf. on Computer Vision 2. Los Alamitos, 2003. 734–741. [doi: 10.1109/ICCV.2003.1238422]
- [12] Mikolajczyk K, Schmid C, Zisserman A. Human detection based on a probabilistic assembly of robust part detectors. In: Proc. of the 8th European Conf. on Computer Vision. 2004. 69–82. [doi: 10.1007/978-3-540-24670-1\_6]
- [13] Tran D, Forsyth D. Configuration estimates improve pedestrian finding. In: Proc. of the Advances in Neural Information Processing System, Vol.20. Cambridge: MIT Press, 2008. 1529–1536.
- [14] Wu B, Nevatia R. Cluster boosted tree classifier for multi-view, multi-pose object detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Rio de Janeiro, 2007. 1–8. [doi: 10.1109/ICCV.2007.4409006]
- [15] Jones M, Snow D. Pedestrian detection using boosted features over many frames. In: Proc. of the 19th Int'l Conf. on Pattern Recognition (ICPR). Tampa, 2008. 1–4. [doi: 10.1109/ICPR.2008.4761703]
- [16] Tian G, Qi FH. Feature transformation and SVM based hierarchical pedestrian detection with a monocular moving camera. Acta Electronica Sinica, 2008,36(5):1024–1028 (in Chinese with English abstract).
- [17] Dalal N, Triggs B, Schmid C. Human detection using oriented histograms of flow and appearance. In: Proc. of the European Conf. on Computer Vision, Vol.2. 2006. 7–13. [doi: 10.1007/11744047\_33]
- [18] Wojek C, Walk S, Schiele B. Multi-Cue onboard pedestrian detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2009. 794–801. [doi: 10.1109/CVPR.2009.5206638]
- [19] Dalal N. Finding people in images and videos [Ph.D. Thesis]. Institute National Polytechnique de Grenoble, 2006.
- [20] Lin Z, Davis LS. A pose-invariant descriptor for human detection and segmentation. In: Proc. of the European Conf. on Computer Vision. Marseille, 2008. [doi: 10.1007/978-3-540-88693-8\_31]

- [21] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(11):1222–1239. [doi: 10.1109/34.969114]
- [22] Li SZ. Markov random field models in computer vision. In: Proc. of the 3th IEEE Conf. on Computer Vision and Pattern Recognition, Vol.2. Washington, 1994. [doi: 10.1007/BFb0028368]
- [23] Rihan J, Kohli P, Torr PHS. Objcut for face detection. In: Proc. of the Indian Conf. on Computer Vision, Graphics and Image Processing. 2006. 576–584.
- [24] <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/caviar.htm>

附中文参考文献:

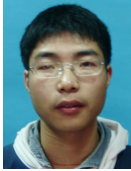
- [16] 田广,戚飞虎.移动摄像机环境下基于特征变换和 SVM 的分级行人检测算法.电子学报,2008,36(5):1024–1028.



郭立君(1970—),男,辽宁凌源人,博士,副教授,主要研究领域为模式识别,计算机视觉.



赵杰焜(1965—),男,博士,教授,博士生导师,CCF 会员,主要研究领域为计算智能,模式识别,人机自然交互.



刘曦(1985—),男,博士,主要研究领域为计算机视觉,图像检索.



史忠植(1941—),男,研究员,博士生导师,CCF 高级会员,主要研究领域为分布式人工智能,机器学习,认知科学.